

LOS REPOSITORIOS DE INFORMACIÓN, GUARDIANES DE LA MEMORIA DIGITAL¹

*Blanca Rodríguez Bravo**

Área de Biblioteconomía y Documentación. Facultad de Filosofía y Letras. Universidad de León.

Resumen: Se estudia el papel de los archivos como guardianes de la memoria digital y productores de información web. Se aborda el reto que los repositorios digitales constituyen en la conservación de los recursos digitales, dando un repaso a los modelos y principales proyectos desarrollados, principalmente en el entorno de las Bibliotecas Nacionales.

Palabras clave: Archivos; bibliotecas; documentos digitales; información digital; repositorios digitales.

Title: THE ROLE OF INFORMATION REPOSITORIES AS DIGITAL MEMORY GATEKEEPERS.

Abstract: It is studied the role of the archives as guardians of the digital memory and creators of information in the web. It is analyzed the challenge that the digital repositories constitute in the conservation of the digital resources. A review to the models and developed projects, mainly in the surroundings of the National Libraries is approached.

Keywords: Archives; digital materials; digital information; libraries; repositories; web resources.

1. INTRODUCCIÓN

Resulta evidente que los archivos han profundizado en la aplicación de las tecnologías de la información a un ritmo espectacular durante los últimos veinticinco años. Sin duda ha influido notablemente la aparición de normativa internacional para la descripción archivística –ISAD(G) e Isaar(CPF)- y para la transmisión de esta información en formatos estructurados como las definiciones de tipo de documento EAD y EAC. Igualmente, la importancia de la gestión archivística en cualquier institución puede apreciarse en la publicación de normas como la ISO 15489 *Information and documentation-records management* (2001) o la ISO 23081 *Information and documentation-records management processes-metadata for records* (2004).

Sin apartarse de las funciones tradicionales de conservadores de la memoria histórica, los archivos desempeñan un papel fundamental en todo lo relacionado con la gestión integral de cualquier tipo de información durante todo su ciclo de vida, desde su nacimiento hasta su expurgo o conservación permanente, convirtiéndose en piezas clave de las instituciones para las que trabajan. Esta consideración aumentará en los próximos años, a medida que estas instituciones difundan a través de la web las ingentes cantidades de información contenidas en los documentos primarios que atesoran.

¹ Texto de la conferencia impartida en las *II Jornadas Técnicas de Archivos en la Administración Local*, organizadas por el Ayuntamiento de Málaga el 9 y 10 de marzo de 2006.

* dphbrb@unileon.es

Esta difusión, sin embargo, planteará, como ya lo ha hecho en otros entornos como el bibliotecario o el académico, problemas para su localización, búsqueda y recuperación en el entorno distribuido de la web actual, y mucho más en la futura web semántica. Estos retos han de solventarse con la creación de un nuevo tipo de estructuras de información y aplicaciones informáticas que no sólo sean capaces de solucionar eficazmente los procesos archivísticos tradicionales, sino de añadir nuevas funcionalidades basadas fundamentalmente en el acceso y la difusión inteligente de la información. De esta forma será posible crear/producir, y no sólo conservar/consumir nuevos recursos electrónicos que estén al servicio de todos los usuarios y del archivo mismo.

2. LOS ARCHIVOS, GUARDIANES DE LA MEMORIA DIGITAL

La mera existencia del documento digital, producido por organizaciones como reflejo de sus actividades, implica la necesidad de que los archivos se planteen su conservación y organización. De las características de estos documentos (compuestos, distribuidos y dinámicos), se deduce su inherente complejidad, incrementada si nos referimos al documento que circula por Internet.

El documento electrónico se compone de señales digitales y, por tanto, carece de los atributos físicos del documento tradicional. Sus características propias, que incluirían la forma o el tipo de material cuando es visualizado en pantalla o impreso, son en gran medida una función del software y están separados del contenido informativo o del contexto del documento.

Durante las dos últimas décadas, los archiveros se han enfrentado a los cambios que representa la tecnología electrónica y sus nuevos documentos que parecen requerir modificaciones en los principios archivísticos tradicionales. En los últimos años, se ha desarrollado una corriente de investigación que propone la sustitución del concepto tradicional "ciclo de vida de los documentos" por el de "flujo continuo de registros", construido sobre la creencia de que la gestión del documento es un proceso continuo desde el momento de su creación.

Existe consenso profesional amplio en la inadecuación de considerar las cuestiones archivísticas al final del ciclo de vida de la documentación. La gestión de documentos y archivos, que tradicionalmente ha sido tratada a posteriori, debe redirigirse a las fases de diseño de políticas, sistemas y aplicaciones. Por ejemplo, es necesario realizar a priori una valoración funcional de los registros que les permita ser autodescritos y autogestionados por el propio sistema. Las aplicaciones desarrolladas hasta el momento utilizan para ello modelos de metadatos, basados en la norma SGML/XML.

El primer paso en una estrategia de preservación de los documentos informáticos de una organización es identificar cuáles son los documentos administrativos que deben ser conservados. Esta tarea no consiste tan sólo en distinguir los documentos con valores primarios, sino también en decidir qué se considera documento dentro del conjunto del sistema informático de la organización.

El dilema actual, ante la complejidad del documento digital, es si la gestión debe orientarse al documento como entidad unitaria, o si debe seguir un enfoque orientado a procesos. El caso paradigmático son las bases de datos, las cuales se pueden concebir como documentos en su conjunto, o únicamente como herramientas generadoras de documentos.

En el primer caso, será preciso intervenir en el análisis y el diseño de la base de datos para garantizar que documenta todas las transacciones y mantiene registros históricos cuando sea necesario. En el segundo caso, los esfuerzos se concentrarán en garantizar que los documentos generados por la base de datos se creen de forma segura y contengan los datos contextuales necesarios.

Un problema similar aparece con el correo electrónico. El sistema de correo electrónico que se utiliza debe permitir discriminar los mensajes con valor de evidencia de una actividad de los mensajes con valor únicamente informativo. Esta funcionalidad no existe en la mayoría de sistemas de mensajería, concebidos únicamente como herramientas de trabajo en grupo. Puesto que el buzón de correo electrónico no está habilitado tampoco para la conservación a largo plazo, es necesario establecer un mecanismo para transferir los mensajes al sistema de gestión documental de la organización, conservando su contexto de transmisión.

Asimismo, toda institución que utilice Internet para sus actividades está dando origen, en algún servidor de la red, a un fondo documental. La conservación y el mantenimiento de este fondo es responsabilidad de la propia institución generadora.

Sin embargo, constatamos que no es frecuente la creación de los correspondientes archivos digitales. Este desinterés es, sin duda, resultado de una falta de conciencia de que los documentos generados por una organización son, sea cual sea su soporte, documentos con valores administrativos, afectados por unos períodos de conservación y por unas especificaciones de acceso. Hasta ahora ha predominado el uso de Internet como medio de difusión, por esta razón circulaban preferentemente documentos cognitivos, sin valor administrativo. Sólo recientemente se está imponiendo como medio de negocio, de formación, etc. La utilización de Internet como medio de negocio se basa en el intercambio de documentación (en este caso claramente administrativa) por vía telemática y en soporte digital, y esta documentación debe ser tratada con una determinada metodología.

Las organizaciones están obligadas a establecer una estrategia para el archivo de sus páginas web por diversas razones, porque la información publicada en una página web es probable que no tenga ningún soporte documental alternativo, ni que exista en otro lugar, porque es importante para una organización saber qué estaba disponible para sus usuarios y cuándo lo estaba, y porque la página web constituye el contexto de utilización de los documentos informáticos que se ofrecen en ella.

La localización física de los documentos informáticos es aleatoria, de modo que la clasificación es la única forma de relacionar dichos documentos. Los documentos digitales no pueden ser conservados como tales más que a partir de la clasificación, que permite relacionar intelectualmente las distintas partes del documento. Esta clasificación debe ser automática, a partir del cuadro de clasificación corporativo introducido en las reglas de procedimiento del sistema.

El segundo paso para facilitar la localización de los documentos informáticos es incorporarles la descripción de su contenido, estructura y contexto. No obstante, puesto que el soporte físico se puede disociar del documento informático, para incluir la descripción no basta con identificar el soporte, debemos añadir la descripción al código del documento y para ello se utilizarán metadatos.

Mediante los metadatos se puede recoger el contexto completo de creación y de utilización de los documentos informáticos, que incluye:

- El contexto administrativo o de creación (autor, fecha, etc.).

- El contexto documental (documentos relacionados, expediente, registros, etc.).
- El contexto tecnológico o de recuperación (formato, versión, programa, etc.).

Los metadatos se pueden generar de forma automática, constituyendo una garantía de la integridad del documento. Por esta razón, establecer una política de metadatos es el primer paso para controlar la producción de documentos informáticos e implica escoger los metadatos que responden a las necesidades de una descripción orientada a la conservación permanente.

A partir de la norma internacional de descripción archivística ISAD(G), la Universidad de Berkeley ha desarrollado EAD, un conjunto de metaetiquetas (DTD) que permiten incluir una descripción archivística normalizada en los documentos. EAD, por tanto, es un estándar para codificar instrumentos de descripción archivística que se basa en la semántica de SGML y XML.

Todas estas propuestas, eminentemente tecnológicas, exigen una actuación previa en la fase de creación de los documentos. A fin de preservar sus valores primarios, el documento debe ser creado de forma controlada, de manera que el sistema garantice su fiabilidad y su autenticidad, y debe ser creado en el formato que mejor responda a su futura disposición, sea ésta la conservación o la destrucción. De este modo, la herramienta conceptual que se convierte en la clave para la preservación de los documentos electrónicos son las reglas de evaluación. Las reglas de evaluación parten de una herramienta previa, un cuadro de clasificación que debe haber identificado y referenciado todos los procesos de una empresa o institución susceptibles de ser documentados.

3. LOS ARCHIVOS, PRODUCTORES DE DOCUMENTOS PARA LA WEB

De acuerdo con la conveniencia de crear contenidos accesibles en la web a partir de la información conservada en archivos, debemos enfrentarnos a las limitaciones que desde el punto funcional presentan los buscadores para encontrar recursos digitales en general, y archivísticos en particular².

Hoy por hoy, las bases de datos permanecen inaccesibles a ellos, lo cual representa un enorme reto a la hora de aumentar la visibilidad de los recursos digitales que producen los archivos.

La solución que está fomentando la comunidad internacional se basa en la recolección de metadatos (*metadatos harvesting*), especialmente en el protocolo OAI-PMH (*Open Archives Initiative-Protocol for Metadata Harvesting*)³.

Éste debe ser un protocolo a implantar en aquellos archivos e instituciones de memoria, que quieran no sólo crear contenidos para la web sino también implantar aplicaciones que favorezcan su localización. Si bien OAI-PMH surgió en el seno de la comunidad académica y científica para la búsqueda y recuperación de documentos científicos en formato electrónico, es aplicable en otros contextos, incluido el archivístico.

Las características básicas de OAI-PMH son:

- Flexibilidad para adaptarse a cualquier ámbito: puede ofrecer información sobre todo tipo de recurso, tanto físico como digital.

² En gran medida este apartado se fundamenta en el artículo de Acuña y Agenjo citado en la bibliografía.

³ <<http://www.openarchives.org/OAI/openarchivesprotocol.html>>

<<http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>>.

- Gran facilidad de implantación. Los componentes del modelo OAI están basados en un recolector de metadatos o *harvester*, en una interfaz de búsqueda y recuperación y en un repositorio común formado por la recolección de repositorios individuales, así como en un conjunto de peticiones y respuestas que se realizan a través de http.

Su estructura es perfectamente inteligible para un profesional de la información y asumible por éste para impulsar el desarrollo de los sistemas de recuperación basados en este protocolo. Muestra de que OAI no es sólo un protocolo para la búsqueda y recuperación de artículos científicos o de material bibliográfico en la web es que puede transmitir y presentar la información en varios formatos, entre ellos EAD.

Son ya millones los recursos digitales accesibles en la red mediante OAI en los distintos recolectores que existen. Dentro de las instituciones que ofrecen sus bases de datos en forma de repositorios OAI recolectables, se encuentran muestras de cómo los archivos pueden aplicar este protocolo y difundir sus colecciones, sirva de ejemplo, *Archives in London and the M25 Area*⁴.

Existe, sin embargo, una inmensa cantidad de información todavía no digitalizada y es necesario subsanar esta situación para edificar repositorios y recolectores basados en OAI. Es preciso que los archivos se animen a crear contenidos para la web actual y a facilitar su difusión en la misma. Ahora bien, para potenciar esta difusión será preciso afrontar la construcción de la web semántica que permitirá promocionar la visibilidad de los recursos digitales existentes y mejorar su búsqueda y recuperación.

La web semántica es un nuevo modelo de estructuración, interrelación y recuperación de la información, gracias al uso de una serie de normas y estándares ya existentes, como XML o RDF⁵, o algunas nuevas creadas al efecto como *OWL Ontology WebLanguage*⁶, así como determinados protocolos de comunicación y estructura de la información que configuran una solución informática, completamente nueva, original y sumamente potente.

Lo que se pretende con la web semántica es superar las carencias e insuficiencias que la *World Wide Web* histórica ha ido presentando a medida que crecía exponencialmente el número de sitios en la Red. La web semántica tiene como objetivo sustentar los sitios y servicios de la web en unas nuevas estructuras de información constituidas por una red organizada de conceptos y relaciones semánticas pertenecientes a un dominio del conocimiento a las que se denomina ontologías. La suma de conceptualizaciones de los diferentes dominios del conocimiento concretadas en ontologías, junto con los agentes de software intermediarios son sus elementos constitutivos, de igual modo que el conjunto de los sitios actuales componen la web que conocemos.

Es en este entorno en el que tienen que desplegar su actividad los archivos. Ni que decir tiene que a medida que estas instituciones se incorporen a la web difundiendo sus bases de datos, sus metadatos y sus documentos digitales, participarán plenamente de esta problemática, de ahí que les interesen las soluciones actualmente planteadas.

⁴ <<http://www.aim25.ac.uk/index.stm>>.

⁵ <<http://www.w3.org/RDF/>>.

⁶ <<http://www.w3.org/TR/owl-features/>>.

Dado que el nacimiento de las ontologías y de la web semántica es reciente, son muchas las interpretaciones y diseños que se discuten, ahora bien, si la contraponemos con la web actual, se pueden establecer las siguientes características distintivas:

- Los recursos no están relacionados únicamente mediante enlaces simples, sino a través de relaciones que tienen un significado.
- A diferencia del tradicional modelo web, el contenido de las páginas, previamente estructurado, es inteligible no sólo para el usuario sino también para las aplicaciones, programas y agentes del software. El entramado de relaciones semánticas lo proporcionan las ontologías y por medio de ellas se puede navegar a significados más amplios, más restringidos y relacionados.
- A través de la estructura semántica que componen las ontologías, tanto los usuarios como los agentes de software recuperan información verdaderamente pertinente. Los contenidos de las bases de datos pasarán a ser recuperables.

El desarrollo de la web semántica dependerá de la disponibilidad de contenidos y del marcado sintáctico y semántico de estos contenidos, así como del desarrollo de herramientas que permitan el análisis de la información. Con este fin se deberán adoptar lenguajes de representación del conocimiento, como RDF u OWL, para lo cual las instituciones de la memoria se encuentran en una situación privilegiada dado que tradicionalmente estructuran la información de que disponen en instrumentos tales como inventarios, catálogos, clasificaciones, tesauros y todo tipo de taxonomías. Muchos de estos instrumentos se encuentran ya automatizados y generalmente en estructuras normalizadas. No obstante son escasos los archivos que han utilizado lenguajes de marcado como XML para estructurar esta información y más aún los que emplean metadatos para atribuir significados mediante una codificación precisa.

Sin embargo, pasar de ISAD(G) a EAD no entraña especial dificultad. Los responsables de archivos tienen que conocer y tomar la oportuna decisión para realizar las transformaciones al nuevo etiquetado. El siguiente paso que se propone es extraer los conceptos y relaciones semánticas de los metadatos y crear ontologías como procedimiento para estar presentes en la futura web semántica. Todo ello habrá de realizarse basándose en las correspondientes aplicaciones informáticas y participando en el diseño de las mismas.

4. LA CONSERVACIÓN DE LOS RECURSOS ELECTRÓNICOS: LA CREACIÓN DE REPOSITARIOS DIGITALES

Es evidente que las bibliotecas, los archivos y los museos han demostrado su capacidad a lo largo de los siglos de preservar los materiales del pasado. Las profesiones a que dieron vida estas instituciones han establecido normas, criterios, pautas, etc., para guiar las políticas y las acciones en cuanto a la preservación de la memoria intelectual de nuestra civilización universal. Ahora se plantea el gran reto de cómo proteger los recursos digitales, tal y como se ha podido hacer con los materiales en soportes tradicionales.

Las administraciones de diversos países han entendido la necesidad de pasar a la acción, y de establecer políticas y emprender acciones de preservación para asegurar la pervivencia de la producción digital, como ya se había hecho históricamente con los documentos impresos y en soportes tradicionales, mediante las leyes nacionales del depósito legal. En la mayor parte de los casos que se mencionarán ha sido la biblioteca nacional

quien ha liderado el proceso de preservación y el acceso al patrimonio digital; para hacerlo ha implicado al resto de agentes. Los archivos han de seguir su ejemplo.

Como señala Alice Keefer (2005b) las dificultades son notables. Para empezar, los métodos tradicionales de preservación de la producción bibliográfica (como el depósito legal) son de difícil aplicación en el entorno digital porque, los recursos digitales pueden instalarse en servidores de cualquier lugar del mundo. En segundo lugar, la producción digital tiene un crecimiento exponencial, siendo además muy variable la durabilidad de los materiales publicados en Internet y, en consecuencia, limitada la posibilidad de acceso permanente al patrimonio. Finalmente, es preciso señalar la cuestión de la propiedad intelectual del producto digital, sin un derecho basado en el principio de copia para la preservación que asegure la conservación y perdurabilidad del patrimonio digital, con las limitaciones comerciales que sean necesarias.

El problema tecnológico, aún sin ser lo más problemático, suele ser el más inmediato y visible. Por ejemplo, lo que más distingue la preservación de materiales digitales de los tradicionales son aspectos como:

- Su dependencia del entorno informático.
- Velocidad de los cambios de este entorno.
- Fragilidad de los soportes.
- Facilidad de modificar los datos (y por lo tanto, la dificultad de garantizar la autenticidad y la integridad de los documentos).
- La intervención activa necesaria a lo largo de la vida del recurso.

Las principales estrategias aplicadas actualmente para frenar la pérdida de información digital son, de forma resumida:

- Preservación de la tecnología.
- Migración de los datos.
- Emulación de las aplicaciones informáticas originales.

Sin embargo, estas estrategias se consideran soluciones a corto plazo a un problema a largo plazo. Es decir, nuestros conocimientos actuales más punteros no nos garantizan la capacidad de preservar lo que estamos creando para un futuro medio y lejano.

La razón por la que nuestros esfuerzos para preservar la memoria digital no son suficientes es que otros retos más allá de los tecnológicos representan obstáculos aún más complejos que superar. Como señala Keefer (2005b) se trata de problemas de distinto tipo:

- Legales: ¿cómo preservar un recurso reproduciéndolo o reformateándolo si no tenemos el permiso del titular?
- Económicos: ¿cómo mantener y actualizar la infraestructura y los recursos necesarios para garantizar la perdurabilidad de los documentos a lo largo de los años?
- Institucionales: ¿cómo asegurar el compromiso institucional permanente?, ¿cómo convencer a todos los implicados de la necesidad de colaborar?, ¿cómo verificar si se cumple todo lo que se promete?

En el último año muchas instituciones se han comprometido a crear repositorios digitales. En cuanto a repositorios institucionales, son más de 300 los registrados en la *Open*

*Archives Initiative*⁷. Característica común a todos estos repositorios es la de estar creados con paquetes de software con licencia *Open Source* y desarrollados por grupos de trabajo vinculados de una manera u otra a la comunidad *Open Archives Initiative*, basada en la implementación de un protocolo común: OAI-PMH.

El principal enfoque de estos depósitos actualmente es asegurar la captación y la descripción de estos recursos para su recuperación posterior, principalmente de acuerdo con la iniciativa de *Open Access*⁸. Pero son pocas las instituciones que han afrontado plenamente el problema de la preservación a largo plazo.

La preservación de los recursos digitales es un reto al que tarde o temprano tendrá que enfrentarse toda institución –biblioteca, archivo, museo– que pretenda preservar para futuros usuarios lo que se genera hoy en formato digital. Sin lugar a duda, hay un largo camino por delante.

4.1 Repositorios digitales nacionales: principales modelos

Un repositorio digital nacional tiene la misión de garantizar el acceso a largo plazo a los recursos digitales que se generan en un territorio, o sobre un territorio determinado.

LLueca ha hallado referencias de veinte casos de depósitos nacionales. Las experiencias existentes se agrupan en dos modelos. Por un lado, el modelo integral o exhaustivo (modelo mayoritario, y característico de Suecia, Noruega, Finlandia, Islandia y Austria, entre otros) que apuesta por la integración automática del total de la web objeto de preservación a partir de determinados criterios infraestructurales (lingüísticos, según el dominio de las webs, según la ubicación del servidor, etc.). Por otro lado, el modelo selectivo (seguido por Australia, Canadá, Japón y el Reino Unido entre otros países) que dirige las acciones de recopilación de acuerdo con una política selectiva temática sobre un espacio geográfico determinado, sobre un tema de interés nacional, etc., mediante acuerdos con los editores o productores de recursos web⁹.

Estos dos modelos han dado paso en algunos países, y cada vez con más fuerza, a modelos híbridos que complementan la captura periódica del total de la web nacional con acuerdos con los productores, que se basan en intereses temáticos o que tienen relación con acontecimientos de actualidad (elecciones, catástrofes, etc.): Dinamarca, Francia y Nueva Zelanda trabajan en esta línea.

La mayor parte de los depósitos que siguen un modelo integral han adoptado medidas para incluir determinados recursos (como publicaciones periódicas) que los acercan a parámetros híbridos. Esta es la tendencia generalizada.

Tres ejemplos que representan los modelos anunciados son: integral (Kulturarw3 de Suecia) <<http://www.kb.se/kw3/ENG>>, selectivo (Pandora de Australia) <<http://Pandora.nla.gov.au/index.html>> e híbrido (Netarkivet de Dinamarca) <<http://netarchive.dk/index-en.php>>.

⁷ <<http://www.openarchives.org>>.

⁸ Budapest Open Access Initiative (BOAI). Open Society Institute (OSI) <<http://www.soros.org/openaccess>>.

⁹ En gran medida este apartado se fundamenta en el artículo de C. Lueca Fonollosa citado en la bibliografía.

4.2 Organizaciones y proyectos suprainstitucionales

Los proyectos mencionados, que llevan a cabo las bibliotecas nacionales, se encuentran a menudo bajo “paraguas” más amplios de cooperación entre bibliotecas u otro tipo de instituciones.

International Internet Preservation Consortium

La organización que agrupa a la mayor parte de estas iniciativas es el IIPC (International Internet Preservation Consortium, <<http://netpreserve.org>>), que tiene la misión de adquirir, preservar y hacer accesible el conocimiento y la información sobre Internet para las futuras generaciones de todo el mundo, promoviendo el intercambio global y las relaciones internacionales.

Fue creado formalmente en julio de 2003 y está formado por 12 miembros.

El IIPC tiene los siguientes objetivos:

- Recoger una parte rica del contenido de Internet de todo el mundo, para ser preservada de forma que pueda ser archivada y asegurado el acceso en el tiempo.
- Fomentar el desarrollo y el uso de las herramientas comunes, técnicas y estándares que permitan la creación de archivos internacionales.
- Animar y ayudar a las bibliotecas nacionales de todo el mundo para archivar y preservar Internet.

Así pues, el consorcio no captura webs, sino que agrupa a una serie de instituciones que lo hacen, y tiene como objetivo promover estas actividades.

Nordic Web Archive

El NWA (Nordic Web Archive, <<http://nwa.nb.no>>) es un foro de las bibliotecas nacionales escandinavas (Dinamarca, Finlandia, Islandia, Noruega y Suecia) para la coordinación y el intercambio de experiencias en los campos de la captura y el almacenaje de documentos web.

Internet Archive

El Internet Archive <<http://www.archive.org>> es una organización sin ánimo de lucro fundada en 1996 para construir una “biblioteca de Internet” y ofrecer acceso permanente a investigadores, historiadores, personal académico y al público en general a las colecciones históricas en formato digital. Se encuentra situado en la antigua prisión de San Francisco (EEUU).

Recibe el apoyo de diversos organismos como la Library of Congress, los US National Archives y los UK National Archives, entre otros.

Actualmente Internet Archive se considera el mayor archivo web del mundo, e incluye texto, audio, imagen en movimiento y software, así como páginas web archivadas de todo el mundo. De acceso abierto y en línea, el gigante contiene un total aproximado de 600 millones de sedes web y cada dos meses realiza una captura masiva que afecta a millones de páginas web, siguiendo el modelo exhaustivo que Suecia y otros países representan.

Recientemente y con sede en Ámsterdam se ha creado el European Digital Archive, rama europea del Internet Archive.

4.3 Repositorios españoles

En nuestro país, existen alrededor de 15 repositorios *Open Access*, muchos de ellos en fase de diseño, y principalmente de carácter institucional, iniciativas de las universidades en su mayoría y de finalidad divulgativa de la investigación.

Un repositorio con finalidad conservadora es PADICAT, el repositorio del Patrimonio Digital de Catalunya, iniciativa sustentada por la Biblioteca de dicha comunidad. Desconocemos si existe alguna iniciativa en marcha en el mundo de los archivos en España.

4.4 Repositorios de colecciones archivísticas

Cabe destacar en este sentido el trabajo desarrollado por la Oficina de Registros Públicos del Reino Unido, *Public Record Office (PRO)* <<http://www.pro.gov.uk/>>, <<http://www.nationalarchives.gov.uk/default.htm>>, cuya finalidad es prestar asistencia para el estudio del pasado y promoverlo seleccionando, preservando y proporcionando acceso a los registros públicos. Dos de los principales proyectos en materia de preservación digital de la PRO son: EROS (*Electronic Records from Office Systems*) y NDAD (*National Digital Archive of Datasets* del Reino Unido). Reseñable resulta, igualmente, el proyecto ya citado *Archives in London and the M25* <<http://www.aim25.ac.uk/index.stm>>, dirigido a proporcionar acceso electrónico a las bases de datos archivísticas y a las colecciones de manuscritos contenidas en los archivos de las universidades del área londinense, un total de 49 instituciones.

5. CONCLUSIONES

Es preciso animar a que los archivos, y también el resto de instituciones informativas, encaminen su actividad, en el marco de las funciones que le son propias, hacia el entorno de la creación de contenidos digitales y su difusión en la web por los mecanismos y tendencias más actuales de la misma. En definitiva se trataría de:

- Hacer accesible internacionalmente el patrimonio cultural español.
- Generar contenidos en los idiomas de España para la web.
- Implementar sistemas de información adaptados a la normativa de la comunidad internacional, vinculada a todo tipo de instituciones de memoria y a las tendencias de las tecnologías de la información en la web.
- Conocer, mediante el desarrollo y uso de las correspondientes aplicaciones y sistemas de información como la recolección de metadatos o la creación de ontologías, el dominio Patrimonio Cultural para contribuir a su difusión, mantenimiento y estudio.

Lo que se propone es que la difusión de la información contenida en los archivos y los objetos digitales asociados (bien sea por digitalización o por tratarse de documentos electrónicos de origen), incorporen siempre sus metadatos conforme a la normativa expuesta para que sea posible el desarrollo de repositorios abiertos en los que descargar sistemáticamente los registros. De esta forma, podrán ser recogidos por recolectores desarrollados por terceros y ofrecer a los usuarios el compendio de información producida por múltiples instituciones.

Dichos repositorios presuponen un cambio de orientación en las políticas conservadoras individualistas de los archivos. La cooperación se impone si se persigue la eficacia.

Los proyectos de repositorios son económicamente costosos y pasan forzosamente por la implicación del número más elevado de agentes posibles que doten de continuidad a los programas una vez iniciados. En este sentido, los fracasos que regularmente afectan a los proyectos mencionados se deben a la falta de financiación.

El interés por la preservación digital se haya generalizado en los países desarrollados, pese a que con un grado de desarrollo heterogéneo. El futuro es híbrido: la diferenciación en modelos (integral *versus* selectivo) representa sólo la primera fase de desarrollo de los proyectos.

Los archivos tendrán que sumar sus esfuerzos y coordinarse en el establecimiento de repositorios digitales que les permita conservar y dar acceso a la documentación administrativa que custodian. El modelo a seguir ya se encuentra formalizado. Se trata de poner manos a la obra.

6. BIBLIOGRAFÍA

- ACUÑA, M. J. DE y AGENJO, X. Archivos en la era digital: problema (y solución) de los recursos electrónicos. *El Profesional de la información*, 2005, vol. 14, nº 6, nov.-dic., p. 407-413.
- BALLEGOOIE, M. V. Metadata for archival collections: the University of Toronto's "Barren Lands" Project. *RLG DigiNews*, vol. 5, nº 4, 2001. <<http://www.rlg.org/preserv/diginews/diginews5-4.htm>>. [Consultado: 7 de mayo de 2005].
- BARRUECO CRUZ, J. M. y SUBIRATS COLL, I. Open archives initiative. Protocol for metadata harvesting (OAI-PMH): descripción, funciones y aplicaciones de un protocolo. *El profesional de la información*, 2003, vol. 12, nº 2, p. 99-106.
- BEARMAN, D. A. Item level control and electronic recordkeeping. *Archives & Museum Informatics*, 1996, vol. 10, nº 3, p. 214-217.
- BEARMAN, D. A. y HEDSTROM, M. Reinventing archives for electronics records: alternative service delivery options. En: HEDSTROM, M., ed. *Electronic records management program strategies*. Pittsburgh, PA: Archives & Museum Informatics, 1993, p. 123-156.
- COLE, T. W. Using OAI: innovations in the sharing of information. *Library hi tech*, 2003, vol. 21, nº 2, p. 115-117.
- COOK, T. What is past is prologue: a history of archival ideas since 1898, and the future paradigm shift. *Archivaria*, 1997, vol. 43, nº 2, p. 17-63.
- COX, R. J. The record in the information age: a progress report on research. *The Records & Retrieval Report*, 1996, vol. 12, nº 1, p. 1-16.
- DOLLAR, C. M. *Archivistica e informatica: l'impatto delle tecnologie dell'informazione sui principi e sui metodi dell'Archivistica*. Macerata: Pubblicazioni della Università, 1992.
- GILLILAND-SWETLAND, A. J. Enduring Paradigm, New Opportunities: the Value of the Archival Perspective in the Digital Environment. En: CLIR Reports, 2000. <<http://www.clir.org/pubs/reports/pub89/contents.html>>.
- HEDSTROM, M. Electronic archives integrity and access in the network environment. *American Archivist*, 1995, vol. 58, nº 3, p. 312-324.

- INTERNATIONAL COUNCIL ON ARCHIVES. Committee on electronic and other current records (CER). <<http://www.ica.org/body.php?pbodycode=CER&plangue=eng>>. [Consultado: 30 de abril de 2005].
- KEEFER, A. Aproximació al moviment "open access". *BID: textos universitaris de biblioteconomia i documentació*, nº 15, diciembre 2005a, <<http://www.ub.es/bid/15keefer.htm>>.
- KEEFER, A. Preservación digital y depósitos institucionales. *El Profesional de la información*, 2005b, v. 14, nº 6, nov.-dic., p. 404-406.
- LAGOZE, C. y SOMPEL, H. V. The making of the Open Archives Initiative Protocol for Metadata Harvesting. *Library hi tech*, 2003, vol. 21, nº 2, p. 118-128.
- LIBRARY OF CONGRESS. Encoded Archival Description <<http://lcweb.loc.gov/ead/>>. [Consultado: 3 de mayo de 2005].
- LLUECA FONOLLOSA, C. Webs siempre accesibles: las bibliotecas nacionales y los depósitos digitales nacionales". *BID: textos universitaris de biblioteconomia i documentació*, diciembre 2005, nº 15, <http://www2.ub.edu/bid/consulta_articulos.php?fichero=15lluec2.htm>.
- MELERO, R. Significado del acceso abierto (open access) a las publicaciones científicas: definición, recursos, copyright e impacto. *El profesional de la información*, 2005, vol. 15, nº 4, p. 255-266.
- NATIONAL LIBRARY OF AUSTRALIA <<http://pandora.nla.gov.au/>>. [Consultado: 13 de mayo de 2005].
- OPEN SOCIETY INSTITUTE (OSI). Budapest Open Access Initiative (BOAI). <<http://www.soros.org/openaccess>>. [Consultado: 10 de mayo de 2005].
- PEIS, E. y FERNÁNDEZ-MOLINA, J. C. Modelos metadatos para la gestión de documentos electrónicos.[CD-ROM]. En: *V Congreso ISKO-España*. Alcalá de Henares: Universidad de Alcalá de Henares, 2001.
- PROM, C. J. Reengineering archival access through the OAI protocols. *Library hi tech*, 2003, vol. 21, nº 2, p. 199-209.
- RODRÍGUEZ BRAVO, B. *El documento entre la tradición y la renovación*. Gijón: Trea, 2002.
- RODRÍGUEZ LÓPEZ, J. Ciencia y comunicación científica: edición digital y otros fundamentos del libre acceso al conocimiento. *El profesional de la Información*, 2005, vol. 14, nº 4, p. 246-254.
- SERRA SERRA, J. Archivar Internet. *El Profesional de la Información*, 1998, 7, nº 7-8, p. 14-18.
- SERRA SERRA, J. Gestión de los documentos digitales: estrategias para su conservación. *El Profesional de la Información*, 2001, 10, nº 9, p. 4-18.
- SHREEVES, S. L.; KACZMAREK, J. S. y COLE, T. W. Harvesting cultural heritage metadata using the OAI Protocol. *Library hi tech*, 2003, vol. 21, nº 2, p. 159-169.
- SIMONS, G.; BIRD, S. Building an Open Language Archives Community on the OAI foundation. *Library hi tech*, 2003, vol. 21, nº 2, p. 210-218.
- SOLBAKK, S. A. Critical technological and architectural choices for access and preservation in a digital library environment. *Library Review*, 2003, vol. 52, nº 6, 2003, p. 251-256.

SOMPEL, H. V.; YOUNG, J. A. y HICKEY, T. B. Using the OAI-PMH... Differently. [En línea]. *D-Lib Magazine*, 2003, vol. 9, nº 7-8. <<http://www.dlib.org/dlib/july03/young/07young.html>>. [Consultado: 26 de noviembre de 2005].

UPWARD, F. The continuum: principles, structures and dualities. *Archives and Manuscript*, 1997, vol. 25, nº 1, p. 27-46.

WARNER, S. E-prints and the Open Archives Initiative. *Library hi tech*, 2003, vol. 21, nº 2, p. 140-150.

7. SELECCIÓN DE ENLACES

Archival Preservation of Smithsonian Web Resources: strategies, Principles, and Best practices. Dollar Consulting, 2001. <<http://www.si.edu/archives/archives/dollar%20report.html>>.

Archiving web resources. Canberra: National Library of Australia, 2004 <<http://www.nla.gov.au/webarchiving/>>.

Archivo Nacional de Australia <<http://naa.gov.au/recordkeeping/preservation/summary.html>>.

Archivos Nacionales de Singapur <<http://www.nhb.gov.sg/NAS>>.

CEDARS project <<http://www.leeds.ac.uk/cedars/>>.

Consejo Internacional de Archivos <<http://www.ica.org/>>.

Conversion and Migration Criteria in Records Keeping Systems. Association for Information Management Professionals. <<http://www.arma.org/standards/index.cfm>>.

DAVID: Digitale Archivering in Vlaamse Instellingen en Diensten (Archivos Digitales de las Instituciones y Administraciones Flamencas). Archivos de la Ciudad de Amberes. <<http://www.antwerpen.be/david/website/eng/index2.htm>>.

DELOS project <<http://www.delos.info>>.

Digital Curation Centre <<http://www.dcc.ac.uk/index>>.

Digital Preservation Coalition <<http://www.dpconline.org/graphics/index.html>>.

Directrices para la preservación del patrimonio digital. Paris: UNESCO, 2003. <<http://unesdoc.unesco.org/images/0013/001300/130071s.pdf>>.

Effective Records Management Project. University of Glasgow. <<http://www.gla.ac.uk/InfoStrat/ERM/>>.

ERPANET project <<http://www.erpanet.org/>>.

Implementing preservation repositories for digital materials: current practice and emerging trends in the cultural heritage community. Dublin Ohio: OCLC, 2004. <<http://www.oclc.org/research/projects/pmwg/surveyreport.pdf>>.

International Records Management trust (IRMT). <<http://www.irmt.org/>>.

National Archives and Records Administration (NARA) (EEUU) <<http://www.archives.gov/>>.

National Digital Information Infrastructure and Preservation Program <<http://www.digitalpreservation.gov/>>.

National Historical Publications and Records Commission (NHPRC) (EEUU) <<http://www.archives.gov/grants/index.html>>.

National Library of Australia. Preserving access to digital information (PADI) <<http://www.nla.gov.au/padi/index.html>>.

Preservation Metadata: Implementation Strategies Working Group (PREMIS)
<<http://www.oclc.org/research/projects/pmwg/>>.

Preserving Cornell's Digital Image Collections: Implementing an Archival Strategy
<<http://www.library.cornell.edu/preservation/IMLS/>>.

Public Record Office (PRO) <<http://www.pro.gov.uk/>>,
<<http://www.nationalarchives.gov.uk/default.htm>>.

Portal de Archivos de la UNESCO <<http://portal.unesco.org>>.

Victorian Electronic Records Strategy Project
<<http://www.prov.vic.gov.au/vers/vers/default.htm>>.